

---

# Determining Ethnic Origin in an Interview Survey

## *Problems and Recommendations*

ACQUISITION OF ACCURATE DATA on the ethnic makeup of the U.S. population is of great concern among numerous Federal agencies. Federal programs provide financial aid to various minority groups, such as Spanish-Americans and American Indians, and the amount of money is relative to the size of the population of the particular group. Undercounts of persons within these groups could result in serious cut-backs in Federal funds.

A number of data collection programs, using many methods, are attempting to obtain a count of persons within each minority group. In 1976, the ongoing household Health Interview Survey (HIS) of the National Center for Health Statistics introduced questions aimed at determining the national origin or ancestry of respondents. These questions were especially designed to identify persons of Spanish origin because they form the largest identifiable minority group in the United States, except for the black population.

When designing the new questions, the Division of Health Interview Statistics personnel looked to the Bureau of the Census for guidance. The Bureau has been collecting information toward identifying the Spanish ethnic makeup of the U.S. population since 1910 (1). In the latest (1970) decennial census, the following ethnic indicators were used: Spanish surname, language spoken in the home as a child, birthplace of parents or of the respondent, and country of origin (2).

In the Current Population Survey (CPS) of the Bureau of the Census, respondents are asked: "Which of the national or ethnic groups on this card best describes your ethnic origin or descent?" They

are shown a flashcard from which they are instructed to select the appropriate group. The ethnic origin card lists the groups as follows:

German	Mexican
Italian	Mexicano
Irish	Puerto Rican
French	Cuban
Polish	Central or South American
Russian	Other Spanish
English	Negro
Scottish	Black
Mexican-American	Another group not listed
Chicano	

The designers of the new questions for the HIS tried several modifications of the decennial census indicators and the CPS flashcard technique before arriving at a final version. In the early stages of experimentation, the HIS obtained information on birthplace of the respondent, birthplace of the parents, childhood language, and origin or descent by using basically the same kind of flashcard used in the CPS.

Analysis of the data from these pretests led to several conclusions. First, the questions on birthplace and parentage had little value because more than 95 percent of the respondents reported the United States as the place of birth for themselves as well as their parents. For the 5 percent not reporting the United States as birthplace, there was no apparent

---

*Tearsheet requests to Claudia S. Moy, Survey Statistician, Division of Health Interview Statistics, National Center for Health Statistics, Rm. 2-44, Center Bldg., 3700 East-West Highway, Hyattsville, Md. 20782.*

difference between place of birth and "origin." Second, the questions on language also had little value as an ethnic indicator. There was a high correlation between persons reporting a Spanish origin and Spanish spoken in the home, but more than half of all persons reporting Spanish spoken did not indicate a Spanish origin, as shown in the following table:

<i>Reported origin</i>	<i>Language spoken as a child</i>		
	<i>Total</i>	<i>Spanish</i>	<i>Other</i>
Spanish .....	47	37	10
Other .....	387	52	335
Total .....	434	89	345

It was also evident that the flashcard list of nationalities created problems because many respondents reported several origins. An entry of two or more origins had to be coded as "multiple," which resulted in the loss of the original information.

The following questions were used on the final HIS questionnaire for 1976 along with Card O to determine the origin of respondents:

Which of these groups best describes \_\_\_\_\_'s national origin or ancestry? If multiple entries: Which of those groups, that is (entries in the preceding question) would you say best describes \_\_\_\_\_'s national origin or ancestry?

*Card O*

- 01 Countries of Central or South America
- 02 Chicano
- 03 Cuban
- 04 Mexican
- 05 Mexicano
- 06 Mexican-American
- 07 Puerto Rican
- 08 Other Spanish
- 09 Other European, such as German, Irish, English, French
- 10 Black, Negro, or Afro-American

- 11 American Indian or Alaskan Native
- 12 Asian or Pacific islander, such as Chinese, Japanese, Korean, Filipino, Samoan
- Another group not listed

A slightly modified version of this procedure is being used in 1977.

**Problems**

At the outset, we recognized that collecting ethnic data in an interview survey would be difficult. No matter how objective the approach is, it deals with a person's self-perception. We could only hope to make the approach understandable to as many people as possible and to keep it consistent with its purpose. A lack of this consistency is a major fault with the questions designed for the HIS. This survey is interested in the respondents' self-identification, that is, to which ethnic or cultural group they believe they belong. And, more specifically, the identification of those persons who believe that they belong to a "Spanish" subpopulation is desired. However, the question, "Which of these best describes your national origin or ancestry?" can only elicit place or line of descent; thus, the respondent's present cultural identity can only be approximated.

Moreover, the questions and flashcard use a mixture of ideas in this attempt to identify the respondent's national origin; countries and continents are mixed with the ideas of race, culture, ethnic identification, nationality, and ancestry. These are not parallel notions—each has a different definition and implies a unique concept to respondents and analysts. The combination of race with nationality is espe-

cially difficult; these are separate items and should be treated separately. On the HIS flashcard, the non-Spanish groups seem basically racially exclusive, whereas the Spanish designators cross racial lines.

The HIS has ranked priorities of minorities by ordering the suggested responses on the flashcard so that they are biased toward the reporting of Spanish groups. Higher counts of Spanish and Spanish-American persons are expected as a result of listing these choices first because it is likely that after the respondent finds an applicable group he will look no further. Thus, a person who is black and also of Spanish ancestry may report only the Spanish origin.

**Survey of interviewers.** Because of the questionable validity of the national origin item, as well as the apparent dissatisfaction among the survey designers, respondents, and interviewers (Bureau of the Census interviewers trained in administering the health survey), information and comments on their experiences with this item were solicited from the interviewers by means of a brief survey. The questions and the interviewers' responses to them follow.

1. Do the respondents seem to have difficulty understanding the concept of national origin or ancestry? How often do you have to probe or explain the meaning of the question? (What kinds of explanations do you usually give?)

Of the 98 interviewers who participated in this survey, 87 (89 percent) indicated having some difficulty in administering the national origin item, and 31 (32 percent) said that they must explain the item to respondents at least half of the time. The kinds of explanation the interviewers gave centered mostly around the nationality or the country from which grandparents or ancestors came before coming to the United States. A few said that they give examples based on their own perceptions of the respondent's background.

2. How often do you decide which category to enter (01-12) because the respondent gives only the name of a country?—50 percent or more of the time, 25-49 percent, 15-24 percent, 5-14 percent, or less than 5 percent.

Thirty-five percent of the 98 interviewers estimated that at least 50 percent of the time they have to decide which code to enter when the respondent does not select a listed category (that is, backcode), and 25 (26 percent) said that this happens less than 5 percent of the time.

3. Do you feel that the respondents are usually knowledgeable enough about their origin to give the correct answer?

4. Do your respondents seem to be more irritated or negative when you ask the question on origin than they are when you ask other questions? If yes, how often does this occur?

More than 78 percent or 77 of the interviewers thought that the respondents are usually knowledgeable about their origin (question 3), and only 19 percent thought that most respondents do not know about their ancestry. Twenty-seven or 28 percent said that respondents seem irritated or negative (question 4), and 13 of these 27 interviewers said that this occurs at least 25 percent of the time. Almost three-fourths (72 percent) said that respondents do not object to the question.

5. What kinds of reactions do the respondents have to the flashcard? (Give examples of comments they make or questions they ask about the card.)

6. Any additional comments you have about the origin question will be appreciated.

7. If you were redesigning this item, how would you word it?

In response to items 5-7, the interviewers offered a number of valuable comments about difficulties encountered. Most indicated that many persons report "American" as their origin, and many of these are confused or upset when they find no suitable category on the flashcard. According to the interviewers, people living in southern States were more likely than others to report "American," or the name of a State, as their national origin or ancestry. Coming from families that have been in this country for many generations, these people appear to have either strong feelings of regional patriotism or no awareness of their country of origin. A few interviewers said that some respondents even consider the question to be un-American.

Because more than 75 percent of the respondents could be expected to be in the "Other European" category, several interviewers suggested that this group be listed first, followed by "Black, Negro, or Afro-American." The eight Spanish identifiers listed first in Card O seem to confuse and anger respondents and also waste time. One interviewer added that some Mexican respondents had difficulty choosing between several of the groups. They were not sure that there really is a difference between Mexican, Mexicano, or Mexican-American, and they did not know which groups would best describe their origin or ancestry.

The category "09-Other European" also is troublesome to respondents. Many wonder "other than what?" Or, not finding their nationality among the examples (German, Irish, English, French), some respondents look for another category containing the rest of Europe. For this reason, many of the interviewers favored expanding the list to include an alphabetical listing of such origins as Dutch, Eastern European, Italian, Middle Eastern, Polish, Russian,

Scandinavian, or Scotch. A few interviewers suggested that all examples be omitted.

Some problems occurred when white persons asked "Where did it say white on the card?" Also, a number of black persons thought that the question was ridiculous because for many of them their group should be obvious. One interviewer also believed that some people reported "American Indian" because they were proud of their Indian heritage, although perhaps they should have reported European as their main origin or ancestry.

It is likely that this short questionnaire sent to the interviewers tended to bias their responses. The assumption that there are problems with the national origin item may be conveyed by the questions on this short survey and hence may be reflected in the interviewers' answers. However, their comments emphasize several major points that need to be considered and corrected. The interviewers seem to be influencing respondents with their own interpretations of the meaning of the item and also with their own perceptions of the origin of the respondents. They may influence respondents further by their attitudes of bewilderment or dislike of the item when these attitudes are present. These problems might be successfully dealt with through additional training of interviewers, providing them with more information about the item, and stressing neutral interviewing techniques.

**Examination of interview data.** So that the actual data from the origin questions could be examined, tables were created from the data in the records of HIS interviews conducted during the first quarter of 1976. The data for the two origin questions were combined to give one "main origin"—either the single entry for the first question on which group best describes national origin or ancestry or the entry for the second question when two or more origins were reported in the first question.

During the first quarter, a total of 148,804,000 persons aged 17 and over were asked the origin question (table 1): 98.1 percent reported a single origin or a combination of origins, and 1.9 percent did not know, refused, or were not asked. About 77 percent of the respondents classified themselves Other European, 10 percent Black, Negro, or Afro-American, 4.7 percent Spanish origin, 1.4 percent American Indian or Alaskan Native, and 1.5 percent Asian or Pacific Islander. The percentage reporting a Spanish origin was comparable to the 4.5 percent reported in the Current Population Report of the Bureau of the Census in March 1972.

Table 1. Main origin of persons aged 17 and over interviewed in the Health Interview Survey, first quarter, 1976

Main origin	Number (in thousands)	Percent
Total .....	148,804	100.0
American .....	3,594	2.4
Central or South American .....	640	0.4
Chicano .....	203	0.1
Cuban .....	605	0.4
Mexican .....	1,168	0.8
Mexicano .....	321	0.2
Mexican-American .....	2,410	1.6
Puerto Rican .....	875	0.6
Other Spanish .....	862	0.6
Other European .....	114,920	77.2
Black, Negro, or Afro-American ..	14,841	10.0
American Indian or Alaskan		
Native .....	2,105	1.4
Asian or Pacific islander .....	2,195	1.5
Other not listed .....	396	0.3
Multiple origin .....	518	0.3
Russian .....	179	0.1
Canadian .....	175	0.1
Unknown, refused, or not reported .....	2,796	1.9

NOTE: Main origin is either the single entry to the question concerning which group best describes national origin or ancestry or a combination of entries if respondents named more than 1 group.

Rather than ask people their race, the HIS interviewers observe the household members present at the time of the interview and infer that the race of the absent members is the same as that of the household respondents. (An eligible respondent is "any 'responsible' adult member of the household . . . mentally competent and physically able to think clearly about the questions . . . 19 years or older or has been married," as defined in the 1976 Health Interview Survey Interviewer's Manual. When children are of racially mixed parents, they are assigned the race of the father, if known. In the remainder of this report, "race" refers to the reported observed or inferred race of the respondent.

The reported national origin is shown in table 2 according to the race of the respondents. Consistency was high between the origin and race for the Spanish, Other European, and Black, Negro, or Afro-American groups. However, discrepancies are evident in the other minorities, which according to HIS rules, should be classified as "other"—not white or black. An overwhelming 84.0 percent of the native Americans were classified by the interviewers as being white or black, with only 16.1 percent classified "correctly." About one-fourth (25.9 percent) of the persons report-

Table 2. Main origin and race of persons aged 17 and over interviewed in the Health Interview Survey, first quarter, 1976 (numbers in thousands)

Main origin	All races	White		Black		Other	
	Number	Number	Percent	Number	Percent	Number	Percent
Total	148,804	134,297	88.2	15,274	10.3	2,232	1.5
American	3,594	3,558	99.0	(1)	(1)	(1)	(1)
Central or South American	640	550	85.9	60	9.4	(1)	(1)
Chicano	203	197	97.0	0	..	(1)	(1)
Cuban	605	596	98.6	(1)	(1)	0	..
Mexican	1,168	1,140	97.6	0	..	(1)	(1)
Mexicano	321	321	100.0	0	..	0	..
Mexican-American	2,410	2,376	98.6	0	..	(1)	(1)
Puerto Rican	875	829	94.8	(1)	(1)	(1)	(1)
Other Spanish	862	764	88.6	74	8.6	(1)	(1)
Other European	114,920	114,831	99.9	66	0.1	(1)	(1)
Black, Negro, or Afro-American	14,841	348	2.3	14,478	97.6	(1)	(1)
American Indian or Alaskan Native	2,105	1,643	78.1	123	5.9	339	16.1
Asian or Pacific islander	2,195	541	24.6	(1)	(1)	1,626	74.1
Other not listed	396	278	70.3	48	12.2	69	17.5
Multiple origin	518	502	96.9	(1)	(1)	0	..
Russian	179	179	100.0	0	..	0	..
Canadian	175	175	100.0	0	..	0	..
Unknown, refused, or not reported	2,796	2,467	88.2	320	11.4	(1)	(1)

<sup>1</sup> Figure does not meet standards of reliability or precision (more than 30 percent relative standard error).

NOTE: Main origin is either the single entry to the question concerning which group best describes national origin or ancestry or a combination of entries if respondents named more than 1 group.

ing Asian or Pacific Islander were observed to be white or black.

A possible explanation for these discrepancies could be that the interviewer frequently "observes" race for the whole household based on the race of one respondent. In racially mixed households, the race of the absent spouse or other household members is recorded as that of the household respondent and is therefore lost. Then, when the respondent gives the origin or ancestry of the absent household members, a difference between race and origin may become apparent. If this problem were a contributing factor to the situation, it seems likely that better correlations between race and origin would exist if data for self-respondents only were studied, self-respondents being defined as persons who responded entirely for themselves or were present (in the same room or within hearing distance) when the first questions were being asked. To check this possibility, table 3 was compiled.

In a comparison of race with origin by respondent status, the preceding premise does not hold. In all cases, the distribution of reported origin by race is about the same for the self-respondents and proxies. If anything, there were more discrepancies among the self-respondents in most origin categories. Therefore, the problem of observing race incorrectly for

mixed households does not appear to be a major cause of race and origin discrepancies.

The large percentage of inconsistencies between race and origin, especially for the minority origins, indicates that either the HIS is making a serious mistake in allowing interviewers to observe race instead of asking the respondents, or that (if we assume that the extent of error in racial classification is not large) race and origin are unique concepts and should not be covered by one term, "national origin." There may be other possible explanations for these poor matches between national origin or ancestry and observed race for these populations groups. Because race is observed by the interviewer, it is subject to the interviewer's attitudes. The origin item reflects the attitudes and interpretations of the respondent. Origin could be related to an ethnic pride—for example, although a person's American Indian ancestry is minor, he may feel proud enough to report it as main origin when the interviewer has "correctly" observed his race as white.

### Recommendations

In view of the results of the first-quarter data analysis and the interviewer comments, a major re-evaluation of the methodology for determining ethnic background seems necessary. The questions and flash-

Table 3. Main origin of persons aged 17 and over interviewed in the Health Interview Survey, by respondent status and race, first quarter, 1976.

Main origin	All races		White		Black		Other	
	Number	Number	Percent	Number	Percent	Number	Percent	
<i>Self-respondents (numbers in thousands)</i>								
Total .....	93,651	82,948	88.6	9,394	10.0	1,310	1.4	
American .....	2,083	2,055	98.7	(1)	(1)	0	..	
Spanish <sup>2</sup> .....	4,255	4,056	95.3	117	2.7	82	1.9	
Other European .....	73,083	73,019	99.9	49	0.1	(1)	(1)	
Black, Negro, or Afro-American ...	9,116	267	2.9	8,849	97.1	0	..	
American Indian or Alaskan Native .	1,406	1,118	79.5	107	7.6	181	12.8	
Asian or Pacific islander .....	1,351	341	25.3	(1)	(1)	990	73.3	
Other <sup>3</sup> .....	896	801	89.4	53	5.9	43	4.8	
Refused, unknown, or not reported .	1,461	1,290	88.3	172	11.7	0	..	
<i>Proxy respondents (numbers in thousands)</i>								
Total .....	55,152	48,349	87.7	5,881	10.7	922	1.7	
American .....	1,510	1,502	99.5	0	..	(1)	(1)	
Spanish <sup>2</sup> .....	2,829	2,718	96.1	51	1.8	61	2.2	
Other European .....	41,837	41,812	99.9	(1)	(1)	(1)	(1)	
Black, Negro, or Afro-American ...	5,724	81	1.4	5,628	98.3	(1)	(1)	
American Indian or Alaskan Native .	699	525	75.0	(1)	(1)	158	22.6	
Asian or Pacific islander .....	845	200	23.7	(1)	(1)	636	75.3	
Other <sup>3</sup> .....	372	335	90.1	(1)	(1)	(1)	(1)	
Refused, unknown, or not reported .	1,335	1,177	88.2	148	11.1	(1)	(1)	

<sup>1</sup> Figure does not meet standards of reliability or precision (more than 30 percent standard error).

<sup>2</sup> Includes Central or South American, Chicano, Cuban, Mexican, Mexicano, Mexican-American, Puerto Rican, and other Spanish.

<sup>3</sup> Includes other not listed, multiple origin, Russian, Canadian, and two origins, did not know which.

NOTE: Main origin is either the single entry to the question concerning which group best describes national origin or ancestry or a combination of entries if respondents named more than 1 group.

card presently being used in the Health Interview Survey are ambiguous, incomplete, and frequently offensive, and as such are not effective for obtaining the required information.

The designing of an effective question to identify minority populations requires several steps. First, the desired groups must be listed and defined. Next, a system for classifying persons within these groups must be devised—that is, a set of identifiable characteristics that would indicate membership in, as well as fit the definition of, each group. The wording of the resulting question should include all characteristics that are intended to be used by the respondent for self-identification. The wording also should be understandable to the respondents (as the interviewers indicated, few persons understood the term “origin,” and only by rewording the question to use common expressions such as “where your grandparents came from,” were the interviewers able to get responses). If a flashcard is to be used, and one seems necessary in order to remind respondents of possible answers, it should be complete, inoffensive, unambiguous, self-explanatory, and—above all—all categories must fit the wording of the question.

Also, the groups should be arranged in an order that encourages accurate reporting and saves time.

In an agreement between the Office of Management and Budget, the Office of Civil Rights, the Equal Employment Opportunity Commission, and the General Accounting Office (3), the five population groups for which data should be collected were outlined. These suggested groups (which are only guidelines and are not to be interpreted as mandatory rules) represent a minimum number of groups, and may be subdivided for greater detail.

1. *American Indian or Alaskan Native*—a person having origins in any of the original peoples of North America.
2. *Asian or Pacific islander*—a person having origins in any of the original peoples of the Far East, Southeast Asia, or the Pacific Islands. This area includes, for example, China, Japan, Korea, the Philippine Islands, and Samoa.
3. *Black, not of Hispanic origin*—a person having origins in any of the black racial groups.
4. *Hispanic*—a person of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish culture or origin, regardless of race.
5. *White, not of Hispanic origin*—a person having origins in any of the original peoples of Europe, North Africa, the Middle East, or the Indian Subcontinent.

The preceding guidelines could be adapted easily

for use in an interview survey. The format entails asking one question to obtain racial or cultural data. The question might read: "Which of those groups best describes your racial or ethnic background?" The interviewer should be instructed to explain, if asked, that we are interested in knowing to which groups the respondents consider themselves to belong.

The flashcard might contain the five basic groups listed (with the elimination of the qualifier "not of Hispanic origin" for groups 3 and 5 to avoid a priority ordering of the minorities). To insure that no group is missed, a sixth category could be added: "Another group not listed—specify."

The interviewers could be given the definitions of the five groups—either printed on the questionnaire or on a separate card to be kept with the interview materials—so that they could clarify the categories when questioned.

Respondents should be permitted to choose more than one group. An additional probe might be helpful: "Which of those groups, that is (*entries in the first question*) would you say BEST describes your racial or ethnic background?" Any multiple responses could then be handled on the coding level.

This method should eliminate some of the problems encountered with the present flashcard. While "origin" is a difficult concept, "race" and "ethnic background" should be easier to comprehend, especially when the respondent is aided with clear definitions. The overabundance of "American" respondents should be lessened because most would classify themselves as white. There would be no confusing partial listing of European countries. Also, the further examples provided to the interviewers, such as North Africa, the Middle East, or the Indian Subcontinent, would help clarify the "white" category when questions arise.

This suggested technique is not ideal, however. It is yet another attempt to combine race with ethnic identification—since four of the five groups are basically racial, the fifth group, "Spanish," does not really fit (and some persons may think that we are calling "Spanish or Spanish-American" a race). However, since both race and ethnicity are included in the question and multiple reporting is allowed, this may not be a serious problem.

An alternative method might be to ask two separate questions, one to determine race and the other to determine culture (4). The question and flashcard technique may be used by asking "What is your racial background?" and showing the respondent a flashcard with the following categories:

1. American Indian or Alaskan Native
2. Asian or Pacific islander
3. Black/Negro
4. White/caucasian
5. Another group not listed, specify

To determine Spanish background, a second question follows: "Is your ethnic background Spanish or Spanish-American?" Or, a slightly less biasing version: "Do you consider yourself to be Spanish or Spanish-American or something else?"

Although this alternative method seems to be the most straightforward, it might be rather awkward in practice. Race is a sensitive subject for many people, especially those belonging to one of the minority categories. It may cause hostility in some cases, jeopardizing any succeeding questions; therefore, this method would be best located at the end of the interview. However, the method does have the advantage of allowing a complete count of all races and permitting analysis of the Spanish population by race.

Whether either alternative, the single-question racial and ethnic identification or the dual-question approach, would be effective in the collection of data on the Spanish population cannot be determined without extensive testing. There may be persons considering themselves to be Chicano, Cuban, Mexican, or Puerto Rican who would not call themselves Spanish or Spanish-American as a result of ethnic pride or sheer misunderstanding.

In preparation for the 1978 Health Interview Survey, the techniques suggested here will be tested in various forms. However, we recognize the limitations of a household survey and the difficulty of collecting data on sensitive topics without risking the alienation of other segments of the population. Further research into methods of identifying minorities is necessary before any methodology can be widely accepted.

### References

1. U. S. Commission on Civil Rights: Counting the forgotten: The 1970 census count of persons of Spanish speaking background in the United States. U.S. Government Printing Office, Washington, D.C., 1974, p. 4.
2. U.S. Bureau of the Census: 1970 census of population: supplementary report on persons of Spanish ancestry. No. PC(SI)-30. U.S. Government Printing Office, Washington, D.C., 1973.
3. National Center for Education Statistics: Announcement: Reporting racial/ethnic data to the Office for Civil Rights. Number 6. U.S. Department of Health, Education, and Welfare, Washington, D.C., June 4, 1975.
4. Federal Interagency Committee on Education: FICE report: vol. 2, No. 1. U.S. Department of Health, Education, and Welfare, Office of Education, Washington, D.C., May 1975.